Review of Barandun and Pohl (tc-2022-117): "Central Asia's spatiotemporal glacier response ambiguity due to data inconsistencies and regional simplifications" – The Cryosphere Discussions

General Comments

The manuscript by Barandun and Pohl highlights the inconsistencies of glacier mass balance responses to different meteorological and morphological drivers in Central Asia's high mountain regions. The authors provide a systematic, region-wide spatial and temporal analysis of two glacier mass balance products against three reanalysis/gridded atmospheric products and open source spatial data on glacier characteristics to this end. The authors find that inconsistencies in glacier response were evident across different glacier or regional/mountain sub-divisions, suggesting that extreme care should be given to statements regarding the main drivers in such a region with limited in-situ validation data.

I think that the authors present a valuable perspective on a common problem in glacier response attribution, specifically applied to a sparsely monitored and highly heterogeneous region, where such drivers are used to theorize about past and future patterns of glacier change in the absence of widespread monitoring. I think that the paper is generally well written and I particularly liked the approach to include specific case studies (Analysis I + II) of the discussion section. Unfortunately, I feel that the comparison of the different reanalysis products and the resultant temporal analyses are not overly convincing or perhaps not robust enough in places. I think the manuscript can be of sufficient quality to be published in The Cryosphere, but a few elements certainly need to be addressed first.

- I think the authors need to restructure their objectives to be something they more clearly address: Objective 1 seems to encompass two separate research questions about the drivers and physical processes themselves, and then the degree to which this is simply a result of which data are used to explore the problem. Objective 2 is related to the latter part of the first objective in my opinion and additionally implies that the authors explicitly compared ground truth data for correlation analysis. I'm not convinced that the authors really test enough the limitations of the gridded data in this manuscript (so objective 1, part 1), or at least not from the data/figures presented.

- It is clear that different reanalysis datasets will produce a different correlation with the same mass balance product. While the authors perform a nice assessment of these expected differences, I'm left wondering how comparable the forcings really are in the first place. The authors apply ERA5 (approx.. 36km), HAR30 (30km) and CHELSA (1km, using ERA5 to downscale). While the first two are closer in terms of spatial resolution and the processes they can represent at the surface, CHELSA represents a major effort to include finer scale topographical features, such as windward/leeward precipitation adjustments, even though it likely remains with significant seasonal biases. Accordingly, I would not expect to find a similar relationship to glacier mass balance (spatially or temporally) when comparing to ERA5 or HAR30 (e.g. L286-7, L343-4), which cannot represent precipitation dynamics at the scales relevant to glacier processes (e.g. L401). The authors have given some consideration to these limitations in their discussion section, but I'm not convinced that it is a fair test and fully supports the conclusions about an ambiguity to glacier response in the region.

  I think that redoing analysis with additional datasets would also not necessarily advance the findings of this manuscript and perhaps cloud the main messages. However, I do think that the authors need to address these discrepancies and perhaps even remove CHELSA from the analysis. More importantly, I would like to see some level of comparison between the different products for air temperature and precipitation estimates (especially if CHELSA remains included). What explains the presence or lack of correlations in given subregions, for example? Some supplementary figure(s) would be useful to that end.

- Building on the previous point, I think the authors also need to address any potential circular issues related to the use of ERA-Interim for the derivation of the $MB_{Baradunetal}$ mass balance results and its comparison to the next generation of the ECMWF reanalysis.

- The temporal analyses produce clearer and more interpretable plots for this manuscript (which I liked), but I have concerns about the robustness of these results given their short temporal scale (14 years for correlations using HAR30) and high uncertainties for mass balances given annual data series. While the ideal minimum sample size required can vary based on the type of observations, it has generally been considered that sample sizes less than 30 can produce imprecise correlation estimates. I'm curious why the authors did not consider a few longer term reanalysis/gridded products of similar spatial resolution (e.g. HARv2, ERA5Land, WRF9km) that all provide data since 1980 at least. I guess there are also temporal limitations for the mass balance data period, but this should at least be discussed more clearly.

- Plotting such large datasets can be challenging and I think the authors have done a great job in many respects. I think there are several areas where they might still be improved for clarity and to help the message of the manuscript. I have given some specific comments below.

- As mentioned, the work is generally well written. However, there are several instances where the wording is unclear or with grammatical errors. I have tried to provide some examples of this below, but the authors should carefully check the entire manuscript at the next submission.

<u>**Specific Comments**</u>

L90+ : I think that the authors provide some very valuable and interesting information about the climatic setting in this paragraph, but I do not feel that the discussion reflects enough upon these interesting variations. Building upon my main point about comparisons of the gridded datasets, the authors should present spatial maps of mean (summed) temperature (precipitation) from different products, or better yet, the spatial trends of some of these variables (where are they statistically significant for ERA5/HAR30?). Exact comparisons can be challenging due to differences in spatial scale, but the authors could consider spatially binning the grids as presented, for example, in de Kok et al. (2020).

L185-188: The authors state that snow depth over-estimation in ERA5 renders use of radiation variables “problematic”, and rely upon only temperature and precipitation data. However, there are well reported cold biases for air temperature at high elevation regions (such as those in this manuscript) which also result from this albedo-effect (e.g. Wang et al. 2020).

L192: Given that the authors test the latest products from ECMWF (ERA5) and CHELSA, I’m surprised that they did not also consider the HARv2 product which is openly accessible and available back to 1980 (thus eliminating the shorter temporal focus and affecting the strength of correlations). I think a line should be added somewhere to justify the use of these three datasets specifically and why they are considered comparable, especially given the points from my main comment above.

L205: I’m not convinced that the MOD10CM product at a 5km resolution does much to aid the analysis presented here. Given its spatial resolution, can it reasonably represent anything meaningful at the scale of the glaciers? Please add some additional information.

L215-216 (and 243-245): I do not understand clearly how the authors relate the spatial means of mass balance with trends of meteorological variables. From the trend you will have just one value per gridded pixel, correct? So the authors combine all glaciers in a given binned region (circles in Figure 1) or just the entire region, and regress all the trends in a given variable (say temperature) at the corresponding pixel of each glacier against the mean or STD of mass balance at that glacier? I believe the authors need to more clearly explain this. Unfortunately, Figure 2 does not help much with this interpretation.

L221-222: Please re-write this sentence to improve syntax. It makes sense, but is not well written.

L225-226: This is also a little unclear from the writing. Do the authors mean that they do not mix variables from different products in their analyses (e.g. temperature from ERA5 and precipitation from CHELSA)?

L242: Please redefine here what is meant by these 'hotspots'. I note that this comes from the author's previous published work, but it would be better to clarify explicitly for the reader what is meant by a hotspot of mass balance variability. I interpret that as the areas in Figure 1 where STD is greatest, for example.

L253: The authors consider the ambiguity of the glacier response here as being largely due to different meteorological or mass balance datasets considered, but what about the uncertainties stemming from the annual-series of geodetic mass balances themselves, which are, to the authors own admission (e.g. L156), with less accuracy, and I assume, less precise? Again, it would be ideal to be able to assess the comparison of the different gridded meteorological and mass balance products independently to identify where the largest differences in the datasets lie, not just the correlations/significance.

L264: I think the authors should attempt to signpost some of the key findings a little better in the figures and simplify some of the main figures where possible (see specific figure comments below). I find it a little hard to navigate from text to figure to follow the storyline.

L344: The authors do not consider all available reanalysis datasets or atmospheric data for their analysis. Please rephrase this.

L346: "at the local scale…" Please check the text for many of these small grammatical issues or minor mistakes.

L346-8: This sentence does not make sense to me. Please rephrase to make your point clear.

L350: but they do not prevent this analysis here, the authors have an annual mass balance value for temporal analysis. I think this first paragraph needs to be restructured slightly to highlight what the authors are able to show, which is normally not considered/available.

L369-379: I think the authors should rewrite this segment as their point is not particularly clear upon reading it. Do the authors mean that their HiMAP regions encompass multiple sub-climatic regions and associated atmospheric processes which therefore limits the usefulness of these defined regions for such analyses?

L382: Please define what "rotation" of mass balance gradients means.

L385: reword "already slight uncertainties…" It does not make sense. Small uncertainties? What estimates? Which variables?

L387: How "invisible"? Are simply not resolved by the coarse reanalysis grid scale?

L388-9: Not clear. Do the authors suggest that coarse grid scale products can capture changes in these regions because there is less sub-grid variability in topography and its associated meteorological complexity?

L392: What do the authors mean with "outer orogene margins"?

L399: "remains unclear…"

L406: please provide examples of which physical snow properties that the authors refer to and what variables missing from the analysis might influence the results.

L410: "extent"

L434: Citation requires parenthesis, check formatting.

L489: Check formatting.

L497-8: Related to the general comments, 14 years does not produce an ideal correlation period. Some discussion on this is needed.

L503: The authors cannot evaluate the biases and patterns of reanalysis datasets due to lack of ground data, but should still highlight how the products relate to each other. Although long term in situ observations are rare, and I understand that it is not the goal of the manuscript to identify the 'best' product to use, but I would be interested to see how the different products compare to even short term measurements from sporadic AWS measurements familiar to the authors from previous work (e.g. Abramov Glacier , Golubin Glacier).

Figures

Figure 1: The inset map in the upper right panel should ideally be of High Mountain Asia to provide a better spatial context as to where the study site is. If able, the authors should attempt to neaten the location and intersect of the legend numbering as there is overlap in places.

Figure 4: Please clarify in the caption how variable importance is defined and why this is given as a relative freq. in the plot labels. Also ideally provide indices (a,b,c etc) to aid navigation for the reader from the main text. I don't find a 'short' analysis period as meaningfully different to the long one for the lower panels. Such morphological characteristics (e.g. debris) won't greatly change over 4 years. Consider removing it to streamline and simplify the figure.
In general I find it very hard to interpret the information given in these figures (4,5,6, A1, A2). Any attempt to simplify it or remove parts would aid interpretation.

Figure 5: I think this figure has far too much information and becomes difficult to interpret. Please consider streamlining the most crucial information where possible, perhaps just showing the short periods, or those with MOD10. For upper panels, one has to struggle to align the x axis label and interpret what is the take-home information. legends should be increased in size, but could easily be kept for just one panel for the whole left and right side.

Figure 6: Similar considerations to the above.

Figure 7: These plots are more clear than the previous ones. Perhaps the authors can re-structure the panels to be 3 rows and two columns, so that the size on the page can be enhanced and the reader can more easily interpret the small circles. Why using 'M' in front of the meteorological datasets? I think it would be clearer without. I do not understand from the figure caption the "with or without snow

cover…” as all show scf as a variable importance coloured on the map. Please clarify or correct this. Please use indices for the subplots to help the reader. For Figure 8 as well.

References

de Kok, R. J., Kraaijenbrink, P. D. A., Tuinenburg, O. A., Bonekamp, P. N. J., & Immerzeel, W. W. (2020). Towards understanding the pattern of glacier mass balances in High Mountain Asia using regional climatic modelling. The Cryosphere, 14(9), 3215–3234. https://doi.org/10.5194/tc-14-3215-2020

Wang, X., Tolksdorf, V., Otto, M., & Scherer, D. (2020). WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis. International Journal of Climatology, May, 1–20. https://doi.org/10.1002/joc.6686